KOMPUTASI: JURNAL ILMIAH ILMU KOMPUTER DAN MATEMATIKA
Vol. 20 (1) (2023), 51-55, p-ISSN: 1693-7554, e-ISSN:2654-3990

Sentiment Analysis of Opinions on the Use of Devices in Students Using the Support Vector Machine (SVM) Method

Muhammad Zuhri[1], **Arie Qur'ania**[2], **Mulyati**[3]

[1,2,3]Department of Computer Science, Faculty of Mathematics and Natural Science, Pakuan University, Bogor, West Java, 16143, Indonesia

Abstract

Sentiment Analysis is a field of science in analyzing a sentiment or opinion on a particular object or problem and the opinion can be divided into several purposes (classes) that lead to negative, neutral or positive opinions. Gadgets (gadgets) are human aids in many fields including work, entertainment, communication and information, the use of gadgets themselves encompasses all ages including school students who use gadgets excessively that affect the mental, physical and attitudes of users. Twitter social media is one of the social media that is used by the public in making opinions about the influence of gadgets, especially parents, these opinions are useful for other users in determining the granting of access rights and direction for children, especially students in using gadgets. Opinion classification is needed in making it easier for other users to see whether opinions from the influence of gadgets fall into the negative, neutral or positive classes. The method used in the classification of opinion is Support Vector Machine (SVM). The data used in this study amounted to 1354 taken in 2019 using web scraping techniques on the Twitter site which are then pre-processed so that it can be processed into the program and classified into 3 classes of sentiments, namely negative, neutral and positive sentiments. In finding the average value of accuracy in the distribution of training data and test data using k-fold cross validation of 10-fold produces an average value of 85.3%. Then testing is done to measure the performance of the SVM method using confusion matrix in the percentage of training data and different test data and produces the highest accuracy value of 83.3%.

Keywords: *Confusion Matrix; K-fold Cross Validation; Sentiment analysis; Support Vector Machine; Twitter.*

1. Introduction

Device users (gadgets) continue to increase every day from year to year, especially since there is a lot of gadget abuse that often occurs in society due to a lack of supervision and restrictions on children under the age of 17. In accessing a site on the internet that is not in accordance with the age limit of gadget users, playing social media, playing online games and other things through an internet connection that can be easily obtained anywhere and anytime which often results in displays or advertisements that are not in accordance with the designation for children under age, coupled with excessive use to the point where they forget to use their devices until late at night.

[*]**Corresponding author.** *E-mail address: mulyati@unpak.ac.id*
Received: 10 November 2022, Accepted: 10 January 2023 and available online 31 January 2023
https://doi.org/10.33751/komputasi.v20i1.6558

This has a tendency towards positive, negative or neutral things, many people have problems with young children to teenagers or students who have a tendency to rely heavily on gadgets which can affect the mindset, mentality, and attitudes of these children [1]. Several previous studies have analyzed the effect of gadgets on children or students including those entitled the impact of using gadgets on the social interaction of children aged 5-6 years with the aim of analyzing the impact of using gadgets on children's social interactions [2]-[5]. Subsequent research regarding parents' perceptions of the presence of the digital technology era among early childhood aims to analyze perceptions or opinions from the parents' side of the development of digital technology among early childhood and determine the level of device use using a quantitative approach with random sampling [6]. Subsequent research regarding the importance of the role of parents in the use of gadgets in early childhood aims to analyze how important the role of parents is in supervising or directing children to use gadgets as they should and teaching that the use of gadgets is in a more positive or beneficial direction [7]. From several previous studies, its only analyzed the influence of gadgets on children's social interactions or students. Therefore, in this study an analysis was carried out on opinion or public opinion regarding the use of gadgets in students whether positive or negative. Sentiment analysis is used to find information to produce to support a decision [8]. One of the methods for analyzing sentiment is Machine Learning with the Support Vector Machine (SVM) method. SVM is a model derived from learning theory statistics that will give better results in comparison other methods [9]. The concept of SVM is to find the best hyperplane that functions as a separator for two classes in the input space by maximizing the distance between classes so that SVM can guarantee high generalizability for future data [10]. It is hoped that this sentiment analysis can become a reference in helping the community make decisions about granting rights to children (students) to use devices and directing the use of gadgets in a better and positive direction.

2. Methods

The method used in conducting this research is the System Development Life Cycle (SDLC) or system development life cycle pattern. The SDLC method has quite complete and structured stages so that it becomes a reference in this study. The SDLC steps are shown in Figure 1 This

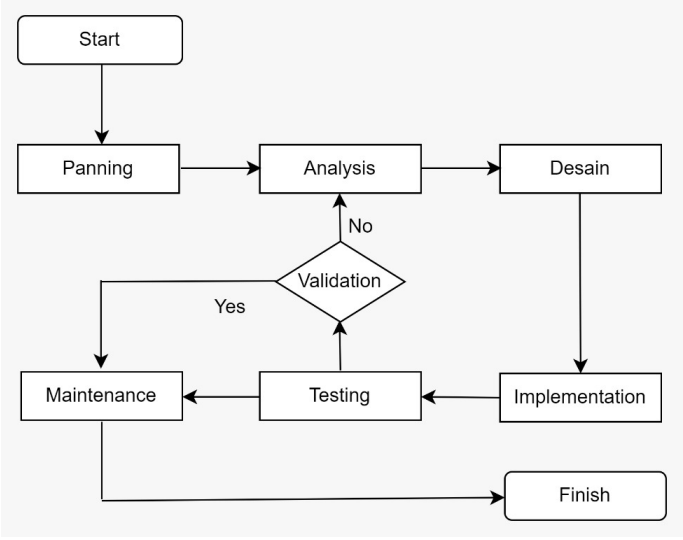

Figure 1. System Development Life Cycle Approach

planning stage has a process that includes two stages in determining research planning, namely Literary Review by examining related or relevant literature sources to conduct this research. The library sources used are in the form of material books obtained from libraries, related theses, and research journals from the internet which are also related to this research. Observation of Secondary Data, namely the data used in this study is indirect (secondary) data obtained from

Twitter using web scraping techniques which are stored in a database in a certain format and then processed for further processing or stages [11].

The system analysis stage is the process of analyzing a system to find the source of the problem and find a solution to the problem by conducting a pattern or flow analysis from start to finish.

This system design stage is a very important process in building a program. In the design of this system there are several stages that need to be carried out, including the design of the system flowchart, input design, process design and output design in the system to be made.

The system implementation stage is the process of applying designs that have been made into a particular programming language to produce an application. Applications will be created using the R programming language with the help of Rstudio software to write programs, create user interface displays and apply the Support Vector Machine (SVM) method.

The System Trial Phase or the testing phase includes validation trials carried out to see the performance of the SVM method using the confusion matrix. The system maintenance stage is when the system has been used until its useful life ends according to the needs of the research being carried out. Some types of maintenance that will be carried out include corrective maintenance, namely system maintenance by making repairs if an error occurs in the system, refinement maintenance, namely maintenance of existing systems by developing or improving and preventive maintenance, namely system maintenance by making system changes. that already exists in its entirety.

The SVM concept can be explained simply as an attempt to find a separator between two classes in the input space. The pattern is a member of two classes: +1 and -1 and various alternative discrimination boundaries. The margin is the distance between the separator and the closest pattern of each class, the pattern that has the closest distance is called a support vector. In solving linear problems, it is assumed that there are training data $\{x_i, y_i\}, x_i$ are attributes for training data $\{x_1, x_2, ..., x_n\}$ and $y_i \in \{-1, 1\}$ are class labels from training data x_i. A good separator can not only separate data but also has a large margin or maximum, data that is near and above the separator [12] . The hyperplane or separator function you are looking for is a linear function using equation 1 as follows

$$f(\overrightarrow{x}) = sign(\overrightarrow{W}^T \overrightarrow{X} + b)\overrightarrow{x} \tag{1}$$

Information:

- $\overrightarrow{W}$: The weight of the hyperplane position in the normal plane
- $\overrightarrow{x}$: Vector data input
- b : Bias

Then for problems in the classification process using Quadratic Programming (QP) which can solve problems with the Lagrange Multiplier so that the classification function becomes like equation 2 as follows:

$$f(\overrightarrow{x}) = sign(\sum_i \alpha_i y_i \overrightarrow{X}_i^T \overrightarrow{X} + b) \tag{2}$$

where α_i is the Lagrange multiplier yang corresponding to $\overrightarrow{x}_i$.

3. Result and Discussion

The data in this study were obtained from text data containing opinions or public opinion on the Twitter social media site related to the influence of gadgets on Kindergarten students to Vocational High School or High School students. The data used is tweet data in the latest Indonesian from January to October 2019 using web scrapping techniques and Google extensions as tools to produce an initial amount of 1415 data which is then pre-processed to become 1354 data. Based on these

data, a test was carried out using the SVM method with the distribution of training data and test data in a structured way to get the best results, and to find the accuracy value of the data that had been processed into the program, a test was carried out using k-fold cross validation in finding the average accuracy of data distribution train and test data. k used is 10-fold or tested 10 (ten) times, where in the first iteration one subset is used as test data and the rest as training data. The test results using the k-fold cross validation that has been carried out can be seen in Table 1 as follows:

Table 1. Result of K-Fold Cross Validation

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
0,838	0,810	0,866	0,866	0,837	0,822	0,889	0,843	0,875	0,888
				average	=	0,853			

From the test results using k-fold cross validation, an average accuracy value of 0.853 or 85.3% is obtained in the partitioning or division of training data and test data and has an error value of 0.147 or 14.7%. Whereas the highest accuracy value of all trials is in the 7th fold with a value of 0.889 or 88.9% and an error value of 0.111 or 11.1%. Furthermore, testing the performance of the SVM method in classifying the sentiment class using a confusion matrix with structured data division to get the best accuracy value. The trial results can be seen in Table 2 below.

Table 2. Result of testing using Confusion Matrix

Trial To-	Training Data	Testing Data	Training Accuracy	Testing Accuracy
1.	948 (70%)	406 (30%)	0,9975	0,830
2.	1015 (75%)	339 (25%)	0,9951	0,826
3.	1083 (80%)	271 (20%)	0,9954	0,833
4.	1151 (85%)	203 (15%)	0,9939	0,822
5.	1218 (90%)	136 (10%)	0,9934	0,830

From the test results in table 1 to find the highest accuracy value based on 1354 data with different percentage partitions for 5 (five) tests, it was found that the third test had the highest accuracy value, namely 0.833 or at least 83.3% of the data successfully classified according to the actual class.

4. Conclusion

The research that has been carried out has produced several conclusions, from trials carried out using 1354 data with several division partitions for training data and test data using k-fold cross validation and producing an average accuracy value of 85.3%. trials were carried out to measure the performance of the SVM method in classifying data in the sentiment class and the highest results were obtained in the third trial out of five tests using the confusion matrix, namely by dividing 1083 training data (80%) and 271 test data (20%) so as to produce an accuracy value of 0.883 or it can be said that of the 271 test data processed as much as 83.3% of the data was successfully classified using the SVM method according to the actual class and from all tests using the k-fold cross validation and confusion matrix it can be said to be quite good because the results of the three trials have different results that are not too significant.

References

[1] Chusna A.P. 2017. Pengaruh Media Gadget Pada Perkembangan Karakter Anak. Jurnal Media Komunikasi Sosial Keagamaan. 17(2), pp. 1-16.

[2] Novitasari .W, Khotimah .N. 2016. Dampak Penggunaan Gadget Terhadap Interaksi Sosial Anak Usia 5-6 Tahun. Jurnal PAUD Teratai. 2(3), pp.182-186.

[3] Doni R.F. 2017. Perilaku Penggunaan Media Sosial Pada Kalangan Remaja. Indonesian Journal on Software Engineering – IJSE. 3(2), pp. 1-9.

[4] Witarsa R. Hadi M.S.R. Nurhananik. Haerani R.N. 2018. Pengaruh Gadget Terhadap Kemampuan Interaksi Sosial Siswa Sekolah Dasar. Jurnal PEDAGOGIK. 6(1), pp. 1-10.

[5] Fitri .S. 2017. Dampak Positif dan Negatif Sosial Media Terhadap Perubahan Sosial Anak. Jurnal Kajian Penelitian Pendidikan dan Pembelajaran. 1(2), pp. 118-123.

[6] Soenarto. Zaini .M . 2019. Persepsi Orang tua Terhadap Hadirnya Era Teknologi Digital di Kalangan Anak Usia Dini. Jurnal Pendidikan Anak Usia Dini. 3(1), pp. 254-264.

[7] Sahriana .N . 2019. Pentingnya Peran Orang tua Dalam Penggunaan Gadget Pada Anak Usia Dini. Jurnal Smart PAUD. 2(1), pp. 60-66.

[8] Afshoh, Fauziah, Pamungkas EW. 2017. Analisa Sentimen Menggunakan Naïve Bayes Untuk Melihat Persepsi Masyarakat Terhadap Kenaikan Harga Jual Rokok Pada Media Sosial Twitter. Skripsi, Program Studi Komunikasi, Fakultas Informatika, Universitas Muhammadiyah, Surakarta.

[9] Darwis D, Pratiwi ES, Octaviansyah EP. 2020. Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia. Jurnal Ilmiah Edutic. 7(1), pp. 1-11.

[10] Suyanto. 2017. Data Mining Untuk Klasifikasi dan Klasterisasi Data, Bandung: Informatika Bandung.

[11] M. Turland. 2010. PHP-Architects Guide to Web Scraping. Marco Tabini Associates.

[12] Kurniawati W. A. 2017. Implementasi Metode Support Vector Machine Untuk Identifikasi Penyakit Daun Tanaman Kubis. Skripsi. Politeknik Negeri Malang.